DOCUMENT RESUME

ED 126 119 TM 005 374

AUTHOR Breland, Hunter M.; Rogosa, David
TITLE Validating Placement Tests.
PUB DATE [Apr 76]
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *College Admission; *College Placement; Content Analysis; Correlation; English; Higher Education; Individual Differences; Multiple Regression Analysis; *Statistical Analysis; Student Placement; *Tests; *Test Validity
IDENTIFIERS Decision Theoretic Utility Model; Trait Treatment Interaction Analysis

ABSTRACT

Recent trends in college admissions are discussed in terms of their influence test validation procedures. In particular, the effects of new college admission practices are considered with respect to the problem of validating placement tests. Traditional test validation techniques are reviewed and compared to the needs specific to placement tests; an example of a placement test validation is presented. Content validity and trait-treatment interaction analyses are stressed in the example analysis, and the possibility of application of decision-theoretic utility models is introduced. In the example analysis, disordinal trait-treatment interactions were found in three colleges. (Author)

VALIDATING PLACEMENT TESTS

Hunter M. Breland

Educational Testing Service

and

David Rogosa

Stanford University

Paper prepared for presentation at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1976.

VALIDATING PLACEMENT TESTS

Hunter M. Breland and David Rogosa[1]

Educational Testing Service / Stanford University

Recent emphases on placement instruments arise from a trend toward the admission into many colleges of students who do not have traditional academic skills. These kinds of admission practices result in a need for special programs suited for handling students of diverse background and preparation. In discussing this problem, Willingham (1974) points to several shifts that have occurred since the late 1950's in the way higher education has adapted to individual differences among students. In the post-Sputnik era there was a fascination with high-level scientific talent, in the mid-1960's there was a shift to more generally selective admissions because of a population bulge in the 18-year-old age group, and in the late 1960's open admission policies came about in response to societal demands. Willingham then notes that:

> More recently it has become clear that access is not enough and that an equally critical problem is how to provide a useful education for students with very different needs and very different backgrounds—i.e., how to deal effectively with wide individual differences that result from free-access policies. From the standpoint of assessing individual differences, the emphasis has changed from identifying students to determining how to educate them. Turnbull (1974) has called it a shift from "which" to "how" (p. 1).

Colleges throughout the country are now experimenting with methods for handling the diversity of entering students. Remedial and compensatory programs, mastery teaching, and personalized systems of instruction represent some of the approaches being tried. Which of the various placement, exemption, and instructional techniques works best, however, is not yet known. But it is clear that placement tests are needed to assist colleges with the instructional problems they face. As new placement instruments are developed, a need arises

[1] Based upon research conducted while the junior author was a Summer Fellow at the Educational Testing Service, Princeton, New Jersey.

to study how they may be used most effectively and how they might be improved. This paper is based upon experiences obtained during studies of an experimental English placement test. The following sections describe methods used, results obtained, and caveats and suggestions for future studies of placement tests.

Content Analyses

If a test is to be used to identify students who have mastered certain material, it is important that it adequately cover relevant topics. A proper placement test will assess a domain of knowledge, skills, and aptitudes that is taught in some specific course of instruction or sequence of instruction. Given a large domain and limited time in which to assess it--which is usually the case--judgments must be made to determine what content is most important and what content is best measured within the confines of a particular form of assessment (e.g., a multiple-choice paper-and-pencil test). To make decisions of this type, committees of subject-matter specialists convene to provide a broad perspective of the domain in question. National surveys are also usually necessary to determine the most equitable representation of topics for a particular test. Following the construction of a test, other groups judge its representation of the domain of interest as well as its appropriateness for particular applications.

To learn what college English teachers thought of the content of the experimental English placement test being studied, questionnaires were sent to 200 English professors in 200 different colleges. The results of this content study showed that, despite considerable controversy within the English teaching profession over what should be taught (as judged from journals and national conventions), there was surprising agreement when professors were asked to rate specified areas of instructional content. The experimental English placement test fared well by this analysis.

Correlational Analyses

Various researchers have questioned the utility of traditional correlational analyses for the study of placement tests (Cronbach, 1971; Hills, 1971; Snow, 1972). Nevertheless, traditions persevere and one may expect that many will continue to consider the correlation coefficient as an important element in any test analysis. Table 1 presents a matrix of correlations relevant to the study of the experimental English placement test (abbreviated as EEPT in Variables 10 and 12 of Table 1). Consideration of the EEPT pretest, Variable 12, shows reasonable correlations with variables the test would be expected to relate to. Note, especially, the correlations of .39 with Fall Grades[2], .43 with an Essay Pretest, .42 with the Essay Posttest graded holistically, .52 with the Essay Posttest graded for grammar, usage, and sentence structure (but abbreviated as simply Grammar in Table 1), and .64 with SAT-Verbal. Observe, also, that the best predictors of the Essay Posttest score (administered in the spring of the freshman year), Variable 4, were Variables 11, 12, and 13, CLEP English Composition, the EEPT Pretest, and the SAT-V Pretest, respectively--all administered at the beginning of or prior to the freshman year of instruction. High School Rank, whether self-reported or college-reported, Variables 9 and 14, tends to have lower correlations with important outcome variables.

While these correlational analyses are interesting--and suggestive of the usefulness of English placement tests like that being studied--temptations to make too much of them should be avoided. Cronbach (1971, p.500) has asserted that "A 'validity coefficient' indicating that test X predicts success within a treatment tells nothing about its usefulness for placement." In the sense that regular freshman English composition is a "treatment," such a view is applicable in the present study. Other writers have espoused a position similar to that of

[2]For short-sequence students, that is, regular freshman English students as contrasted to students placed in a longer sequence for purposes of remediation, compensatory programming, or vertical sectioning.

Table 1

Correlation Matrix for Combined Colleges

Variable Description: Antecedent (A)	Outcome (O)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Fall Grades (Long-sequence)	O	1.00 025	-	.25 370	-	-	-	.52 206	.52 206	.15 617	.41 108	.36 110	.36 921	-	-
2. Fall Grades (Short-sequence)	O	-	1.00 3880	.26 113	.44 94	.57 94	.55 528	-	-	.34 2082	.29 210	.45 453	.39 3879	.32 866	.32 159
3. Essay Pretest	A	.25 370	.26 113	1.00 483	-	-	-	-	-	.21 451	-	-	.43 482	-	-
4. Essay Posttest (Holistic)	O	-	.44 94	-	1.00 96	.48 96	-	-	-	.34 95	-	.47 95	.42 96	.40 91	.24 85
5. Essay Posttest (Grammar)	O	-	.57 94	-	.48 96	1.00 96	-	-	-	.40 95	-	.53 95	.52 96	.29 91	.32 85
6. Spring Grades (Short-sequence)	O	-	.55 528	-	-	-	1.00 528	-	1.00 528	.30 264	-	-	.33 528	.37 78	-
7. Spring Grades (Long-sequence)	O	.52 206	-	-	-	-	-	1.00 860	1.00 860	.23 306	.31 109	.15 109	.24 860	.07 83	-
8. Spring Grades (Total)	O	.52 206	.55 528	-	-	-	1.00 528	1.00 860	1.00 1388	.28 570	.27 131	.14 153	.30 1388	.18 161	-
9. High School Rank (Self-report)	A	.15 617	.34 2082	.21 451	.34 95	.40 95	.30 264	.23 306	.28 570	1.00 3905	.38 119	.46 269	.34 3898	.32 181	.57 166
10. EEPT Posttest	O	.41 108	.29 210	-	-	-	-	.31 109	.27 131	.38 119	1.00 440	.74 408	.70 440	.56 309	-
11. CLEP English Comp.	A	.36 110	.45 453	-	.47 95	.53 95	-	.15 109	.14 153	.46 269	.74 408	1.00 688	.66 688	.68 575	.36 159
12. EEPT Pretest	A	.36 921	.39 3879	.43 482	.42 96	.52 96	.33 528	.24 860	.30 1388	.34 3898	.70 440	.66 688	1.00 7272	.64 1099	.38 167
13. SAT-V	A	-	.32 866	-	.40 91	.29 91	.32 78	-	.18 161	.32 181	.56 309	.68 575	.64 1099	1.00 1099	.33 165
14. High School Rank (College report)	A	-	.32 159	-	.24 85	.32 85	-	-	-	.57 166	-	.36 159	.38 167	.33 165	1.00 167

Note: Correlations above, number of cases below.
Correlations based on less than 50 cases not shown.

Cronbach (e.g., Snow, 1972; Hills, 1971; Thorndike, 1971).

Trait-Treatment Interaction Analyses

Trait-treatment interaction (TTI) analysis is considered by some researchers to be the most useful method for analyzing placement tests (Cronbach, 1971; Willingham, 1974). In the instructional setting, an interaction implies that the advantage of a long-sequence of instruction over a short-sequence of instruction varies according to the level of the placement test score obtained prior to instruction.[3] The notion of TTI is inherent in the logic of placing students with different levels of knowledge in different educational treatments. The question of interest in a placement situation is "Will a student be better off in the normal treatment or in an alternative treatment?" An answer to this question clearly requires information comparing outcomes for a particular placement score for both the conventional and the alternative treatment groups.

The importance and usefulness of the TTI is best understood by examining the regression of the end-of-sequence criterion on the placement test for the two groups of interest: (1) students placed in long-sequence instruction, and (2) students placed in short-sequence instruction. In the optimal case, the regression lines will differ substantially between treatments, as shown in Figure 1. Note that the regression line for the long-sequence group is relatively flat, while the regression line for the short-sequence group is steeper. The advantage of placement into the long-sequence for those on the lower portion of the placement test scale is apparent from an inspection of the differences in the regression lines. Regression line C-D, in Figure 1, represents what might be expected if students were randomly placed (regardless of placement test score) in a special long-sequence of instruction, such as a remedial English

[3]The term, long-sequence, includes both remedial instruction of longer duration than regular (short-sequence) instruction and non-remedial instruction of longer duration than regular instruction.

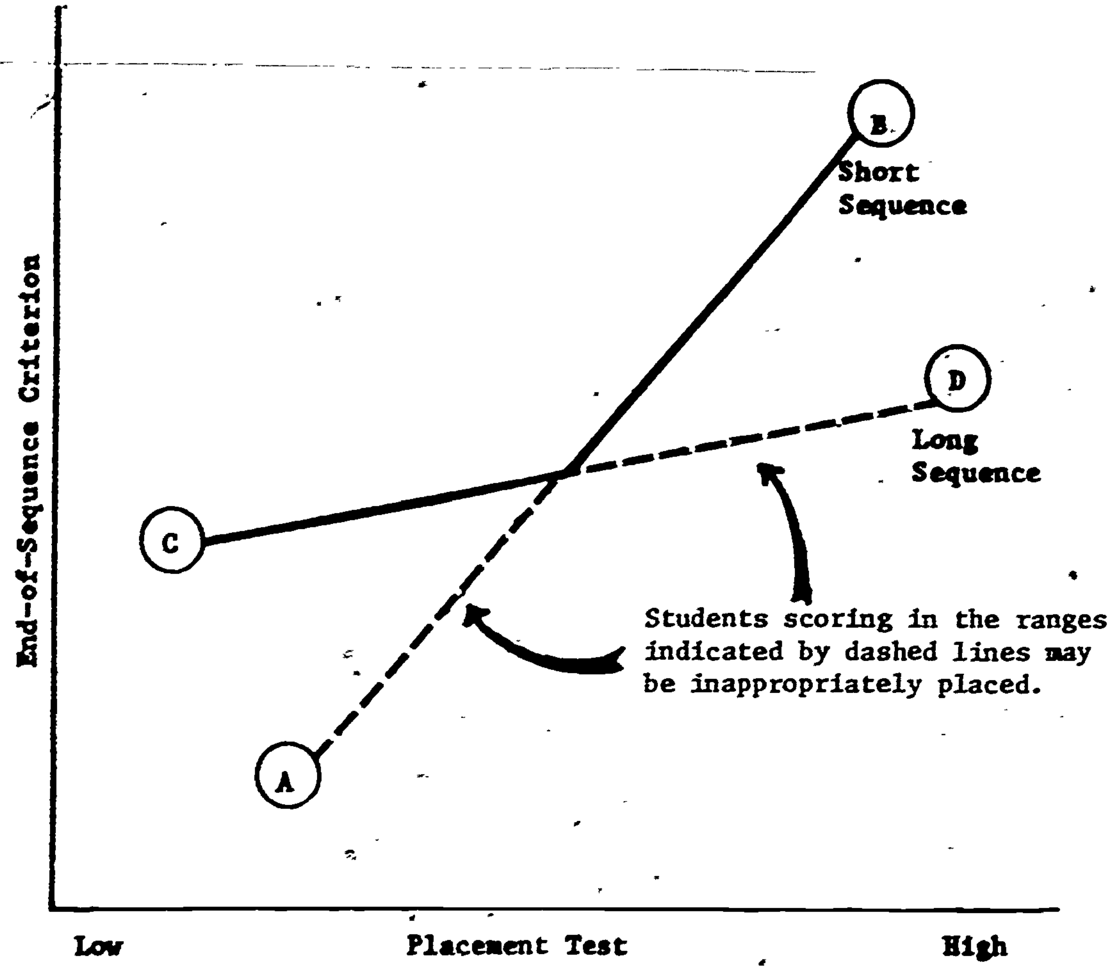

Figure 1.[a] Illustration of the TTI assumption in the case of placement.

[a]Adapted from Willingham (1974)

course of some kind of compensatory programming. The end-of-sequence criterion might be course grades at the completion of a regular course in, say, freshman English. The regular course represents the end of the long-sequence; the remedial course is only the first part of the long-sequence. Therefore, grades in the remedial course are not appropriate for use as a criterion in the TTI analyses. Line C-D shows that students with lower scores on the placement test performed better at the end of instruction than did similar students who were placed in the sho. - (regular) sequence represented by line A-B. The regression lines of Figure 1 are, of course, hypothetical. These kinds of outcomes will not occur unless the placement instrument is finely tuned to the instruction--especially to the remedial instruction.

Methodology for TTI Analysis. The analysis of TTI data is a process of comparing the within-treatment regressions of a suitable criterion variable on the placement test. Non-parallel regression lines indicate that a trait-treatment interaction exists. There are different kinds of interactions, however. Using the language of Cronbach and Gleser (1965), ordinal interactions are indicated by non-parallel lines which do not intersect in the range of interest, whereas disordinal interactions are indicated by lines which intersect in the range of interest. Clearly, disordinal interactions (see Figure 1) are of primary interest for placement decisions. Assuming a valid criterion variable is available, the point of intersection provides a straightforward cutting point for assignment to alternative educational treatments.

The statistical comparison of two regression lines requires that the two groups have similar distributions. One measure of importance is the variance about the regression lines (residual variance); the residual variances should be equal or nearly equal for proper comparison of the regression lines. Large-sample tests of the hypothesis of equal residual variances are provided by Gulliksen and Wilks (1950) and Stroud (1972).

If the residual variances are not significantly different, the next step is the test for equal regression slopes. Establishing a difference in regression slopes is the key evidence in the detection of a trait-treatment interaction. Standard regression theory (e.g., Kendall & Stuart, Vol. II, 1967, p. 371-372) provides a t-test, or an equivalent F-test, for the significance of the difference of the estimated regression slopes. Unfortunately, tests for interaction have relatively little power (Cohen, 1969; Cronbach & Snow, in press). Consequently, failure to reject the null hypothesis of equal regression slopes, when samples are not large, cannot be regarded as conclusive evidence of the absence of TTI. Examination of the estimated regression slopes and the associated confidence intervals supplements the hypothesis testing and provides a more detailed description of the data and of the likelihood of TTI.

Investigating the possibility of a disordinal interaction is the final step in the TTI regression analysis. Graphical inspection is often useful for a rough determination. Statistical inferences can be made for the point of intersection of the regression lines. If the true cutting score (the abscissa of the point of intersection), which we denote as x_o, lies in the range of interest of the predictor variable (the placement test score), then a useful disordinal interaction exists. Robison (1964) demonstrated, under the assumption of equal residual variances, that the maximum likelihood estimator of x_o is

$$\hat{x}_o = \frac{a_1 - a_2}{b_2 - b_1} \quad \text{, where}$$

a_i and b_i are the estimated intercept and slope, respectively, for the regression equation in group i. Kastenbaum (1959) derives confidence intervals for x_o (assuming normality) based on the t-distribution. The width of this confidence interval is a measure of the precision and, therefore, the usefulness of the cutting score. This confidence interval for x_o is identical with the region of nonsignificance obtained from the Johnson-Neyman Technique with one predictor.

Some Results of TTI Analyses

Regression analyses were performed on the schools in the study which provided sufficient data. The variances about the regression lines for the two placement groups for each of the schools were first examined. Most of the standard errors of estimation were quite similar within schools and the null hypothesis of no difference was tenable.

Table 2 shows the number of cases determining each regression line (N), the estimated regression slopes (b), the standard deviations ($\hat{\sigma}(b)$), and the estimated intercept of the regression lines (a). Also shown is the estimated reliability of the experimental test for each group within each school ($r_{xx'}$) along with the regression slopes corrected (b*) for the attenuating effect of measurement error in the EEPT scores. Although the reliabilities between schools vary considerably, the reliabilities between groups within the schools are quite similar. This within-school similarity of estimated reliabilities lends some credibility to the assumption that the t-test for equality of observed regression slopes is a reasonable, albeit approximate, substitute for an exact test for the equality of estimated regression slopes for true scores.

Of course, the stability of the regression coefficients determined depends strongly on the sample size. Consequently, significance tests for differences of regression slopes will have much more power in the schools with large sample sizes. Because of small sample sizes in some of the schools, appreciable differences in the regression slopes will usually fail to be significant. In this exploratory analysis, effects that were not statistically significant were not disregarded, but conclusions from these effects were viewed with appropriate cautions.

Figure 2 illustrates a typical reporting of the results of a TTI analysis of placement test data. For each school the within-group regression lines are plotted. The indicated F-test for differences in slopes is performed to

Table 2

TTI Analysis Statistics

College Code	Long-sequence N	b	∂(b)	a	r_{xx}'	b*	Short-sequence N	b	∂(b)	a	r_{xx}'	b*
A	84	.033	.012	.437	.83	.040	141	.041	.009	.701	.82	.050
B	157	.017	.008	1.77	.76	.022	346	.025	.006	1.56	.78	.032
C	452	.033	.006	.960	.79	.042	711	.052	.004	.166	.72	.072
D	63	.042	.015	.583	.75	.056	49	.017	.010	1.67	.76	.022
E	244	.019	.007	1.85	.73	.026	299	.047	.009	.472	.70	.066

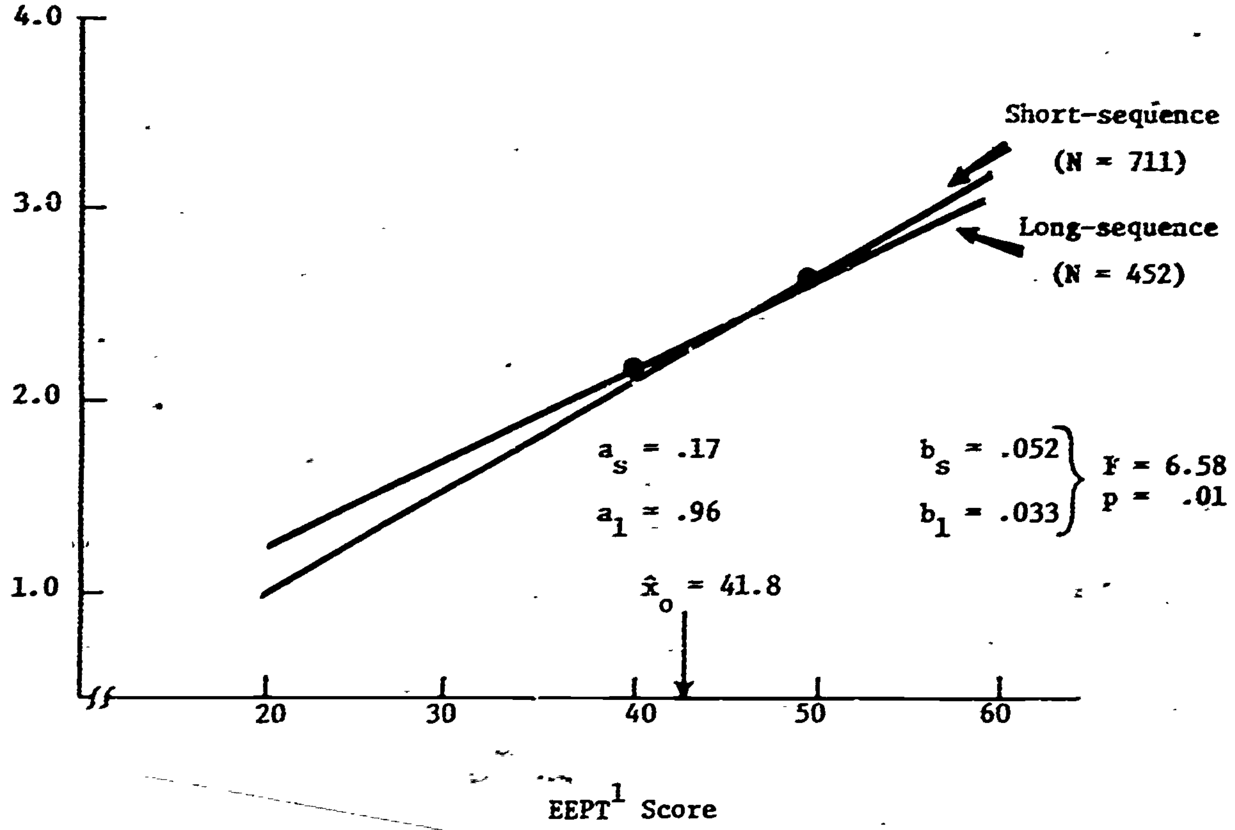

Figure 2. Regression lines for College C.

Notes:

1. An experimental English placement test.
2. The dots indicate the group means.
3. The F statistic indicated is for the hypothesis of equal regression slopes.
4. If b* is used to determine the regression lines, the point of intersection decreases slightly.
5. The symbols, a_s, a_1, b_s, and b_1, represent intercepts (a) and slopes (b) for short (s) and long (l) sequence instructional groups.

establish the TTI effect. Finally, the estimated cutting score $\hat{x}_o$ is shown.

Some Unsolved Problems in TTI Analysis. Implicit in decisions to place students in different instructional treatments is the assumption that the trait used for placement and the treatment interact. In other words, the assumption is that optimum learning, or some other optimum outcome, is maximized by the placement procedure. Implicit in all of these assumptions is one that is perhaps less apparent. This is the assumption that, even if desired educational outcomes are maximized by the placement procedures, there is also an economic or other justification for the placement—for example, that it is worth an extra allocation of financial resources to bring about optimum educational outcomes. Although it is of considerable importance, an analysis incorporating the full decision-theoretic framework, including costs, is beyond the scope of this paper.

Despite the theoretical attractiveness of TTI, in practice it has not often been as useful as hoped. More often than not, interactions of the type desired fail to occur. Reasons why TTI's are difficult to conduct include:

(1) Bias and unreliability in common student performance criteria, such as grades.

(2) Uncontrolled instructional variables. Interactions are most likely to occur when instruction is closely tuned to the test.

(3) Problems specific to the curriculum structure. Willingham (1974) observes that TTI effects will be best seen in a "segmented sequence" of courses (e.g., Mathematics courses). In an "ordered series" of courses, such as in psychology and English curricula, end-of-sequence performance criteria are often insensitive to treatments occurring at earlier stages.

Even when interactions do occur, it is not always certain what interpretation to make. The utility of outcomes is a function not only of student performance but of other factors—including the cost of instruction. It is impossible to construct a single criterion scale integrating even performance plus cost let alone still other important factors such as student satisfaction. Nevertheless, to the degree that local administrators and instructors can incorporate judgmental factors with the TTI analyses, the results can be useful.

From a methodological perspective, the present art of TTI analysis has even further unsolved problems:

(1) Power. As has already been noted, the power of statistical tests is low unless the N is high, but the difficulty of data collections required for TTI analysis tends to reduce the N available.

(2) Measurement Error. Since measurement error in the placement test scores serves to flatten the regression slopes and thereby mask TTI effects, correction for this attenuation is desirable. However, this correction complicates the statistical analysis since the distribution of the corrected slopes is unknown if the reliability is estimated. If the cutting score for the observed regression lines lies near the group means, then it will be little affected by any measurement error correction.

(3) Fixed Predictor Variables. Standard regression theory assumes that the predictor variables are fixed; that is, the observed values are predetermined and replicable from sample to sample. Clearly, this assumption is violated by placement tests. The inferences from the standard analysis are then conditioned on the observed values of the placement test scores, and generalizations to situations with other observed values are not strictly valid. No satisfactory methods for handling these problems of inference exist.

(4) Units of Analysis. In the usual classroom situation, students within a class are not exposed independently to the educational treatment. Therefore, class membership should be taken into account in the examination of treatment outcomes. When students are treated in groups, TTI effects have three possible explanations. They may arise from the individual's response to the treatment, a class effect, or from a comparative effect within a class. The examination of between-class and within-class regressions is helpful in separating these interaction effects. (See Cronbach & Webb (1975) for an illustration of these techniques.) In placement situations each of these three effects can be important for the proper allocation of educational resources.

(5) Choice of Criterion Variable. The outcome measure chosen is crucial to the success of the TTI study. The criterion measure should reflect the instructional objectives and not vary widely over different classes or schools if these are to be pooled in the analysis.

Beyond TTI Analysis

Over the years since Cronbach and Gleser (1965) elaborated possible uses of decision theory in personnel (and other) common decision problems, much has been said and written about the potential of such approaches. Because of theoretical and philosophical issues surrounding applications of decision theory, however, it has been used little in practice. In education, it has not been used at all. A need has existed for simple operational procedures that might embody some of the basic concepts of utility. Davis, Hickman, and Novick (1973); Hambleton and Novick (1973); and Peterson (1974) have described utility models for use in both instruction and selection.

A recent paper by Livingston (1974) describes an operational utility-based approach that may be applicable to the kind of placement situations considered in the present paper. The decision procedure described by Livingston is useful when a decision-maker must take one of two possible actions, say, Accept (A) or Reject (R). Or, the choice may be between Accelerated (A) or Regular (R) instruction. If the test score cut-off point is x_o and the criterion "indifference point" is y_o,

$$u_a\,(y_i) = \text{utility of action A for person } i,$$

$$u_r\,(y_i) = \text{utility of action R for person } i,$$

and

$$u_a\,(y_o) = u_r\,(y_o) = o,$$

then an increasing utility function u_a and a decreasing utility function u_r may be imagined as shown in Figure 3. The utility of the decision procedure is the sum of the utilities of all the individual decisions:

$$U\,(x_o) = \sum_{x_i \geq x_o} u_a\,(y_i) + \sum_{x_i > x_o} u_r\,(y_i).$$

The utility of an ideal procedure is used for comparison (based on knowledge of actual performance):

$$U\,(y_o) = \sum_{y_i \geq y_o} u_a\,(y_i) + \sum_{y_i > y_o} u_r\,(y_i),$$

and a utility ratio,

$$U_r = \frac{U\,(x_o)}{U\,(y_o)},$$

is computed. Note that, unlike correlation and regression coefficients, U_r is a function of both x_o and y_o. Thus, unlike correlation and regression coefficients, U_r is potentially useful to test users not only in evaluating the usefulness of a particular test for their particular purposes, but also in setting cut-off points (provided they can define their indifference point, y_o).

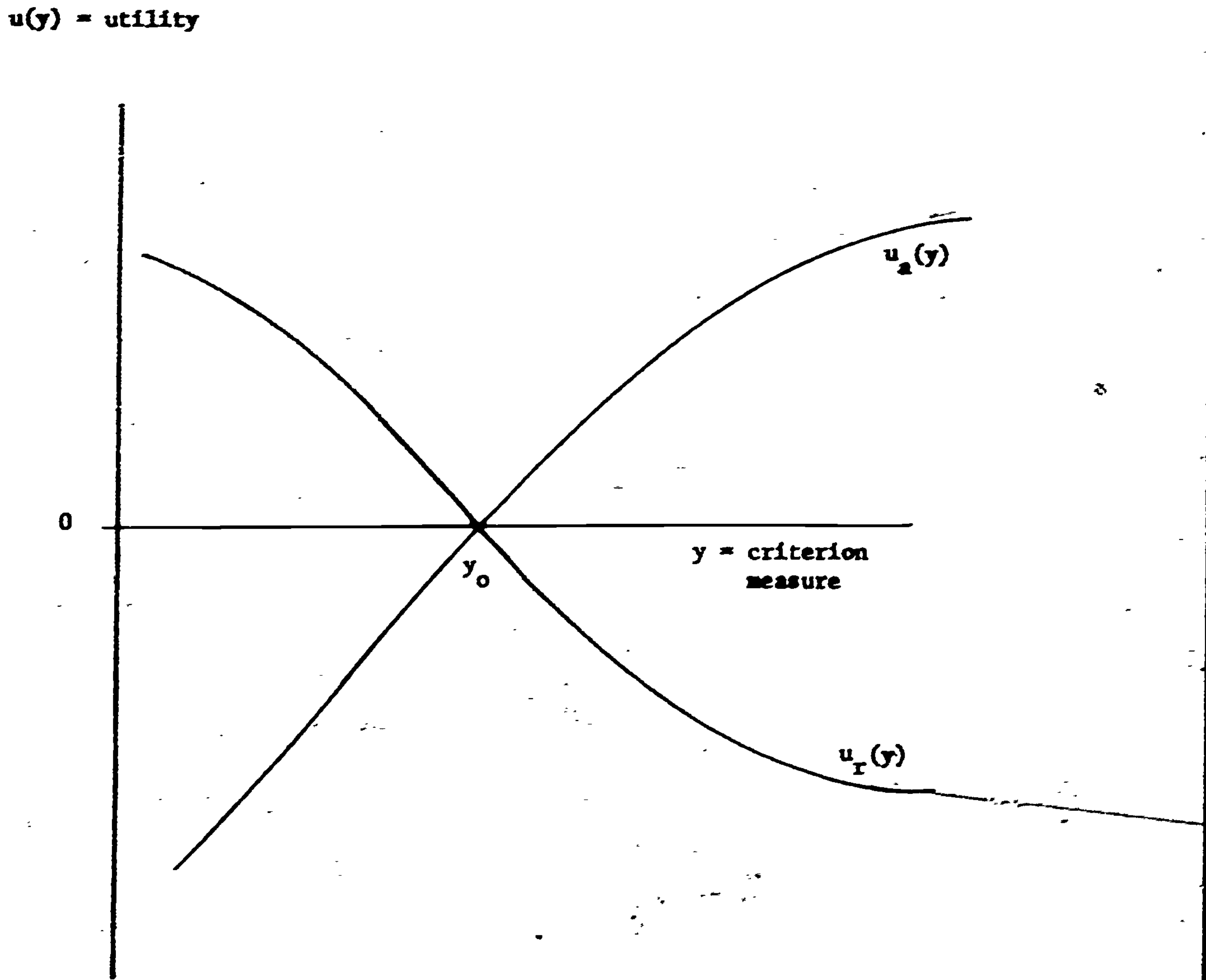

Figure 3. Illustration of one possible choice of utility functions.*

*From Livingston (1975).

There are at least two problems with U_r. First, it would be very difficult for test users to construct utility functions like those in Figure 3. Because of this problem, Livingston suggests that a convention be established, say, to use simple straight-line functions--unless some reason for doing otherwise exists. A second problem with U_r, as described, is that it assumes that there are no constraints on the numbers of persons assigned to A or R. Nevertheless, a decision-maker may be wise to consider ths problem initially without constraints and then to modify the cut-off as suggested by U_r with possible constraints in mind.

References

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1969.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), Educational measurement. Washington: American Council on Education, 1971, p. 443-507.

Cronbach, L. J., & Glesser, G. C. Psychological tests and personnel decisions. (2nd ed.) Urbana: University of Illinois Press, 1965.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods: a handbook for research on interactions. New York: Irvington Press, 1976, in press.

Cronbach, L. J. & Webb, N. Between-class and within-class effects in a reported aptitude X treatment interaction: Reanalysis of a study by G. L. Anderson, Journal of Educational Psychology, 1975, Vol. 67, No. 6, 717-724.

Davis, C. E., Hickman, Jr., and Novick, M. R. A primer on decision analysis for individually prescribed instruction. Technical Bulletin No. 17. Iowa City, Iowa: American College Testing Program, 1973.

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Gulliksen, H., & Wilks, S. S. Regression tests for several samples. Psychometrika, 1950, 15, 91-114.

Hambleton, R. L., and Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Hills, J. R. Use of measurement in selection and placement, in Robert L. Thorndike (Ed.), Educational Measurement. (2nd ed.) Washington, D.C.: American Council on Education, 1971, pp. 680-732.

Kastenbaum, M. A. Confidence interval on the abscissa of the point of intersection of two fitted linear regressions. Biometrics, 1959, 15, 323-4.

Kendall, M. G., and Stuart, A. The advanced theory of statistics, Volume II, New York: Hafner Publishing Company, 1967.

Livingston, S. A. A utility-based approach to the evaluation of pass/fail testing decision procedures, Center for Occupational and Professional Assessment, Educational Testing Service, 1975.

Petersen, N. S. An expected utility model for "optimal" selection. Technical Bulletin No. 24. Iowa City, Iowa: American College Testing Program, 1974.

Robinson, D. E. Estimates of the points of intersection of two polynomial regressions. Journal of the American Statistical Association, 1964, 59, 214-224.

Snow, R. E. Personal/intellectual differences and new forms of education: Aptitude-treatment interactions and individualized alternatives in higher education. Paper presented to the Graduate Records Examination Board Invitational Conference on Cognitive Styles and Creativity in Higher Education, Montreal, Canada, November 10, 1972, 27 pp. Available from the author, School of Education, Stanford University, Stanford, California.

Stroud, T. W. F. Comparing conditional means and variances in a regression model with measurement errors of known variances. Journal of the American Statistical Association, 1972, 67, 407-412.

Thorndike, R. L. Educational measurement for the seventies, in Robert L. Thorndike (Ed.), Educational measurement. (2nd ed.) Washington, D.C.: American Council on Education, 1971, pp. 3-14.

Turnbull, W. W. Testing shifts from "which" to "how". New York Times, Annual Education Review, January 16, 1976, p. 82.

Willingham, W. W. <u>College placement and exemption</u>. New York: College Entrance Examination Board, 1974.